

Figure 1.

```
Sequence type explicitly set to Protein
Sequence format is Pearson
Sequence 1: new_S100_cytokine      104 aa
Sequence 2: G491246                110 aa
Sequence 3: W27152                 98 aa
Start of Pairwise alignments
Aligning...
Sequences (1:2) Aligned. Score: 23
Sequences (1:3) Aligned. Score: 34
Sequences (2:3) Aligned. Score: 29
Start of Multiple Alignment
There are 2 groups
Aligning...
Group 1:                          Delayed
Group 2:                          Delayed
Sequence:1      Score:0
Sequence:3      Score:839
Sequence:2      Score:724
Alignment Score 444
CLUSTAL-Alignment file created [/data4/genetools/lrastelli4630clustalw]
```

Multiple Alignment:

```
new_S100_cytokine MGQCRSANAEDAQEFSDVERAIETLIKNFHQYSVEGG-KETITPSELRLDVTQQLPHLMP
W27152            -----MAAAEPLTELESIEIVVTFFTFARQEGRKDSLVSNEFKELVTQQLPHLLK
G491246           -----MSQLERNIETIINTFHQYSVKLGHPDTLNQGEFKELVRKDLQNFLK

new_S100_cytokine SNCG----LEEKIANIGSCNDSKLEERSEFWELIGEAAKSVKLERP...VRGH...(SEQ ID NO:6)
W27152            DVGS----LDEKMKSLDVNQDSELKFNEYWRLLIGELAKEIRKKKD...LKIRKK...(SEQ ID NO:10)
G491246           KENKNEKVI EHI MEDLDITNADKQLSFEFFIMLMARLITWASHEKMHGDEGPGHHHKPGLG

new_S100_cytokine ----
W27152            ----
G491246           EGTP (SEQ ID NO:11)
```

Figure 2.

Multiple Alignment:

```
new_S100_cytokine      -----DNRTLTKSPDTWS-TMGQCRSANAEDAQEFSDVERAIETLIKNFHQYS
7971c.7__r0s0-212.2__EXT SISSCGAGYRTDDKTQLTEGRTSWP GTMGQCRSANAEDAQEFSDVERAIETLIKNFHKYS

new_S100_cytokine      WEGGKETLTPTS ELRDLVTQQLPHLMPSNCGLEEKIANLGSCNDSKLEFRSFWELIGEAAK
7971c.7__r0s0-212.2__EXT WAGGKETLTPTA ELRDLVTQQLPHLMPSNCGLEEKIANLGN CNDSKLEFRSFWELIGEAAK

new_S100_cytokine      SVKL ERPVRGH (SEQ ID NO:6)
7971c.7__r0s0-212.2__EXT SVKMERPVTRS (SEQ ID NO:3)
```

Figure 3.

hmmpfam - search a single seq against HMM database

HMMER is freely distributed under the GNU General Public License (GPL).

```

-----
HMM file:                pfamHMMs
Sequence file:           /data4/genetools/lrastelli4423Aa315020ProteinFasta.txt
-----
Query:  AA315020

```

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
S_100	S-100/ICaBP type calcium binding domain	40.9	2.9e-08	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
S_100	1/1	32	74	1	44	40.9	2.9e-08

Alignments of top-scoring domains:

S_100: domain 1 of 1, from 32 to 74: score 40.9, E = 2.9e-08

```

      *-->LEkaietiInvFhqYSgreGdkdtLsKkELKeLlekELpnfLkn<--*
          E+aiet+I+ FhqYS  eG k tL+  EL+ L++++Lp+ +
AA315020  32  VERALETLIKNFHQYS-VEGCKETLTPSELRLDLVTQQLPHLMPS  74(SEQ ID NO:33)

```

//

Figure 4A.

BLOCKS Protein Domain Analysis of new_S100_cytokine

Probe Size: 104 Amino Acids

Probe File: lrastelliblocks.seq

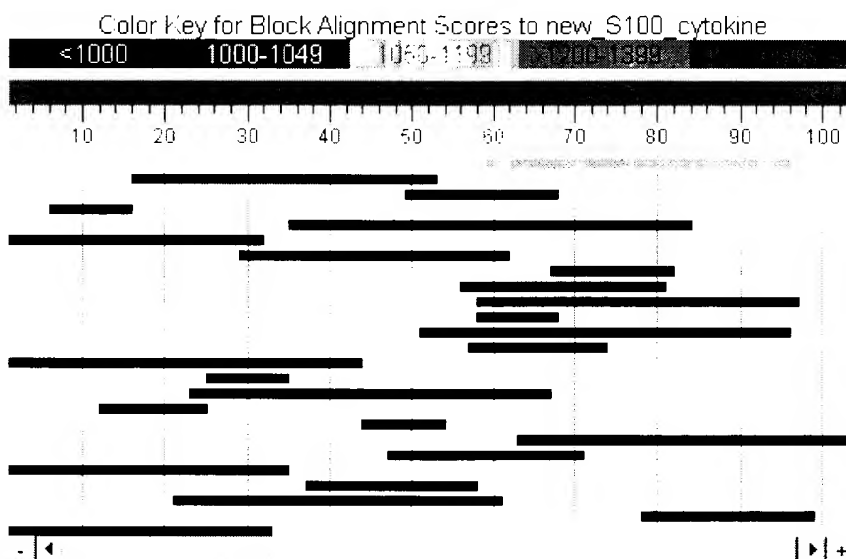
Target File (s) : blocks.dat

Records Searched: 4034

Scores Done: 4034

Alignments Done: 535470

Mouse-over to show define and scores. Click to show alignments



ACC	Description	Strength	Score	IT	ACC	Description
EL000001	" S-100/ICaBP type calcium binding protein.	1.036	1077	0	59	" NCGLLEENKIANLGSCrDCKLFRSFWELIGEAANKV (SEQ ID NO:12)
EL000002	" S-100/ICaBP type calcium binding protein.	1.045	1078	0	16	" DVERAIEETLIKvFHYQSVLEGKvTltpvElrdLvt00 (SEQ ID NO:13)
EL000003	" Bacterial type II secretion system protein F	1.056	1081	0	49	" VTQQLpHLp3ncgLEEKv (SEQ ID NO:14)
EL000004	" Ubiquitin carboxyl-terminal hydrolase family	1.077	991	0	6	" ANAIDAQITv (SEQ ID NO:15)
EL000005	" Bacterial chemotaxis sensory transducers prot	1.058	990	0	35	" GGRKtLtpv1EDVVTQQLpHLp3ncgLEEKv (SEQ ID NO:16)
EL000006	" Phosphoenolpyruvate carboxykinase (ATP) prote	1.017	989	0	0	" mgQcR3ANvAdvQvF3DVERAIEETLIKv (SEQ ID NO:17)
EL000007	" Prokaryotic-type carbonic anhydrases proteins	1.039	987	0	29	" hqF3vegGKvEtltPvElrdLVTQQLpHLp3ncgLEEKv (SEQ ID NO:18)
EL000008	" Ergosterol biosynthesis ERG4/ERG14 family pro	1.099	979	0	67	" IANLGSCrDCKLIEFR (SEQ ID NO:19)
EL000009	" Lysosome-associated membrane glycoproteins du	1.022	976	0	56	" Lp3ncgLEEKvANLGSCrDCKLIEFR (SEQ ID NO:20)
EL000010	" Phosphofructokinase proteins.	1.041	976	0	50	" p3ncGLLEEKIANLGSCrDCKLIEFRvEELIGEAANKV (SEQ ID NO:21)
EL000011	" FH domain proteins profile	990	976	0	54	" P3NCGLLEEKv (SEQ ID NO:22)
EL000012	" Hyptoxins proteins.	1.177	974	0	51	" gqLpHLp3ncgLEEKvANLGSCrDCKLIEFRvEELIGEAANKV (SEQ ID NO:23)
EL000013	" Phosphatidylinositol-specific phospholipase X	1.027	974	0	57	" mP3ncGLLEEKvANLGSCrDCKLIEFR (SEQ ID NO:24)
EL000014	" Glypicans proteins.	1.008	972	0	-1	" mgqGr3ANvAdvQvF3DVERAIEETLIKv (SEQ ID NO:25)
EL000015	" Membrane attack complex components / perforin	1.107	972	0	25	" IANLGSCrDCKLIEFR (SEQ ID NO:26)
EL000016	" Urease nickel ligands proteins.	1.092	970	0	28	" TLIDvthqy3vegGKvTltpvElrdLVTQQLpHLp3ncgLEEKv (SEQ ID NO:27)
EL000017	" Phosphoglycerate mutase family phosphohistadi	1.038	968	0	12	" qvF3DVERAIEETLIKv (SEQ ID NO:28)
EL000018	" Ribosomal protein L29 proteins.	1.092	967	0	44	" ELRLVtqQL (SEQ ID NO:29)
EL000019	" 2'-5'-oligoadenylate synthetases proteins.	1.076	967	0	63	" LeekvANLGSCrDCKLIEFRvEELIGEAANKV (SEQ ID NO:30)
EL000020	" Beta-lactamases class B proteins.	1.052	964	0	47	" dLVTqgLpHLp3ncgLEEKvANLGSCrDCKLIEFR (SEQ ID NO:31)
EL000021	" Glycoprotein hormones beta chain proteins.	1.028	963	0	1	" gQcR3ANvAdvQvF3DVERAIEETLIKv (SEQ ID NO:32)
EL000022	" Vinculin family talin-binding region proteins	1.067	963	0	37	" hqLpHLp3ncgLEEKvANLGSCrDCKLIEFR (SEQ ID NO:33)
EL000023	" Heat shock hsp70 proteins family profile	1.000	960	0	21	" IeTILIDvthqy3vegGKvTltpvElrdLVTQQLpHLp3ncgLEEKv (SEQ ID NO:34)
EL000024	" Hydroxymethylglutaryl-coenzyme A lyase protei	1.008	960	0	78	" Lp3ncGLLEEKvANLGSCrDCKLIEFR (SEQ ID NO:35)
					-0	" HgqGr3ANvAdvQvF3DVERAIEETLIKv (SEQ ID NO:36)

Figure 4B.

	10	20	30	40	50	60	
Table3	MGQCRSANAEDAQEFSDVERAIETLIKNFHKYSVAGKKETLTLPaelRLDLVTQQLPHLMPS					
AA007220	MGQCRSANAEDAQEFSDVERAIETLIKNFHOYSVEGKKETLTLPSELRLDLVTQQLPHLMPS					
Consensus	MGQCRSANAEDAQEFSDVERAIETLIKNFH YSV G KETLTLP ELRLDLVTQQLPHLMPS					
	70	80	90	100			
Table3	NCGLEEKIANLGNcNDSKLEFGsFWELIGEAAKSVKMERPV (SEQ ID NO:3)					
AA007220	NCGLEEKIANLGSNDSKLEFRsFWELIGEAAKSVKLERPV (SEQ ID NO:39)					
Consensus	NCGLEEKIANLGNcNDSKLEF sFWELIGEAAKSVK ERPV (SEQ ID NO:40)					

Figure 4C.

	10	20	30	40	50	60
Table 6					
AA007220	MGQCRSANAEDAQEFSDVERAIETLIKNFHQYSVEGGKETLTPELRDLVTQQLPHLMPS					
Consensus	MGQCRSANAEDAQEFSDVERAIETLIKNFHQYSVEGGKETLTPELRDLVTQQLPHLMPS					
	70	80	90	100		
					
Table 6	NCGLEEKIANLGSCNDSKLEFRSFWELIGEAAKSVKLERPVRGH (SEQ ID NO:6)					
AA007220	NCGLEEKIANLGSCNDSKLEFRSFWELIGEAAKSVKLERPVRGH (SEQ ID NO:39)					
Consensus	NCGLEEKIANLGSCNDSKLEFRSFWELIGEAAKSVKLERPVRGH (SEQ ID NO:40)					

Figure 4D.

	10	20	30	40	
Table 3	ERAIETLIKNFHYSV-AGKKETLTPELRDLVTQQLPHLM	(SEQ ID NO:3)			
gi 4139958 pdb 1MHO	EKAIVVALIDVFHQYSGREGDKHKLKKSELKELINNELSHFL	(SEQ ID NO:41)			
PROTEIN MRP-126	EKAIDVILIDVFHQYSRREGDKDTLTKELKLLIEKQLANYL	(SEQ ID NO:42)			
ICTACALCIN	QKGMALLISTFHKYSGKEGDKCTLTKGELKDLITKELGGAF	(SEQ ID NO:43)			
CALGRANULIN B	ESSIETIINIHFQYSVRLGHYDTLTKKEFKQLVQKELPNFL	(SEQ ID NO:44)			
Consensus	I FHYSG L L E L L	(SEQ ID NO:45)			

Figure 4E.

	10	20	30	40	
Table 6	BRAIETLTKNFHQYSV-EGGKETLTPELRDLVTQQLPHLM (SEQ ID NO:6)			
gi 4139958;pdb 1MHO	EKAVVALIDVFHQYSGREGDKHKLKKSELKELINNELSHFL (SEQ ID NO:41)			
PROTEIN MRP-126	EKAIDVILIDVFHQYSRREGDKDTLTKELKLLIEKQLANYL (SEQ ID NO:42)			
CALGRANULIN B	ESSIETIINIHFQYSVRLGHYDTLTKKEFKQLVQKELPNFL (SEQ ID NO:44)			
CALGRANULIN B	ERSITTIIDTFHQYSRKEGHPDTLSKKEFRQVVEAQLATEM (SEQ ID NO:46)			
Consensus	E I FHQYS G L E L L (SEQ ID NO:47)			

Figure 5

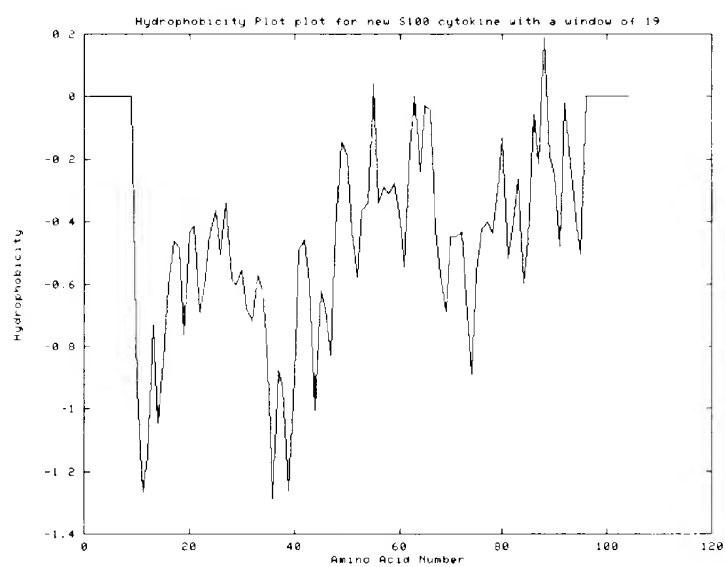


Figure 6

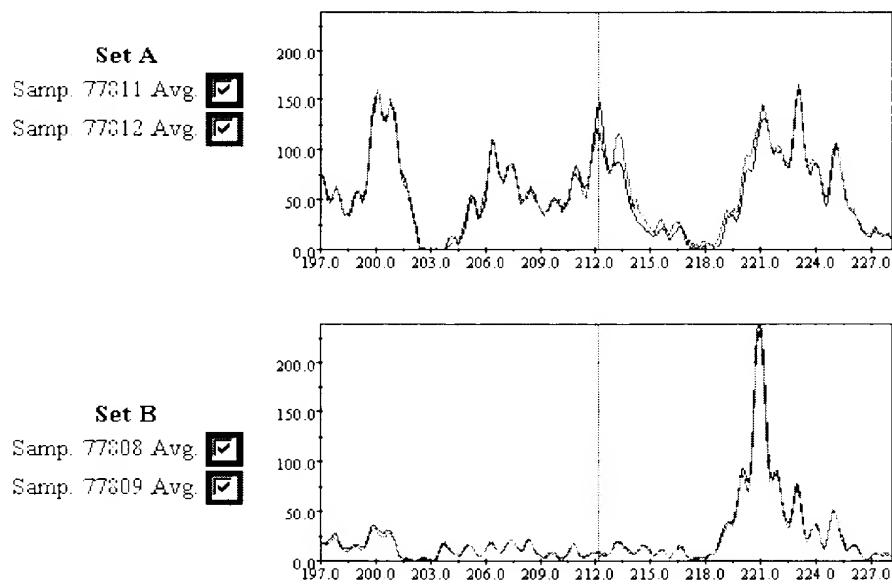


Figure 7

***** Contig 1 *****	
65677221+	GAATTCCAGAGGGAGTTCTCAGTCCCCCGGACAGGCCTCTCCAGCTTCACACTCTTGGC
AA315020-	TGCCCCCGGACAGTCCTCTCNAGCTTCACACTCTTGGC
consensus	GAATTCCAGAGGGAGTTCTCAGTCCCCCGGACAGGCCTCTCCAGCTTCACACTCTTGGC
65677221+	CGCTTCTCCAATCAGCTCCCAGAAACTCCTGAACTCCAGTTTATAGAGTCATTGCAGCTGCC
AA315020-	CGCTTCTCCAATCAGCTCCCAGAAACTCCTGAACTCCAGTTTATAGAGTCATTGCAGCTGCC
consensus	CGCTTCTCCAATCAGCTCCCAGAAACTCCTGAACTCCAGTTTATAGAGTCATTGCAGCTGCC
65677221+	CAGGTTGGCAATTTTCTCTTCCAGGCCACAGTTGCTCGGCATGAGATGGGGCAGCTGCTG
AA315020-	CAGGTTGGCAATTTTCTCTTCCAGGCCANAGTTGCTCGGCATGAGATGGGGCAGCTGCTG
consensus	CAGGTTGGCAATTTTCTCTTCCAGGCCACAGTTGCTCGGCATGAGATGGGGCAGCTGCTG
65677221+	GGTGACCAGGTCCCGTAGCTCAGAAGGGGTCAGCGTCTCCTTCCCACCCTCCACGGAGTA
AA315020-	GGTGACCAGGTCCCGTAGCTCAGAAGGGGTCAGCGTCTCCTTCCCACCCTCCACGGAGTA
consensus	GGTGACCAGGTCCCGTAGCTCAGAAGGGGTCAGCGTCTCCTTCCCACCCTCCACGGAGTA
65677221+	CTGGTGAAAGTTCTTGATGAGGGTCTCAATGGCCCTCTCCACATCACTGAATTTC (SEQ ID NO:37)
AA315020-	CTGGTGAAAGTTCTTGATGAGGGTCTCAATGGCCCTCTCCACATCACTGAATTCTCTGAGC
consensus	CTGGTGAAAGTTCTTGATGAGGGTCTCAATGGCCCTCTCCACATCACTGAATTCTCTGAGC
AA315020-	ATCCTCTGCGTTGGCTGACCGACACTGTCCCATGGTGCTCACTGTGTCTGGTCTTTTGGT
consensus	ATCCTCTGCGTTGGCTGACCGACACTGTCCCATGGTGCTCACTGTGTCTGGTCTTTTGGT
AA315020-	GAGAGTTCTGTTGTCCTAT (SEQ ID NO:4)
consensus	GAGAGTTCTGTTGTCCTAT (SEQ ID NO:5)